

A Method of Homogenizing the Extremes and Mean of Daily Temperature Measurements

P. M. DELLA-MARTA

Climatology and Meteorology Research Group, University of Bern, Bern, Switzerland, and National Climate Centre, Bureau of Meteorology, Melbourne, Australia

H. WANNER

Climatology and Meteorology Research Group, University of Bern, and NCCR Climate, Bern, Switzerland

(Manuscript received 30 June 2005, in final form 20 December 2005)

ABSTRACT

To be confident in the analyses of long-term changes in daily climate extremes, it is necessary for the data to be homogenized because of nonclimatic influences. Here a new method of homogenizing daily temperature data is presented that is capable of adjusting not only the mean of a daily temperature series but also the higher-order moments. This method uses a nonlinear model to estimate the relationship between a candidate station and a highly correlated reference station. The model is built in a homogeneous subperiod before an inhomogeneity and is then used to estimate the observations at the candidate station after the inhomogeneity using observations from the reference series. The differences between the predicted and observed values are binned according to which decile the predicted values fit in the candidate station's observed cumulative distribution function defined using homogeneous daily temperatures before the inhomogeneity. In this way, adjustments for each decile were produced. This method is demonstrated using February daily maximum temperatures measured in Graz, Austria, and an artificial dataset with known inhomogeneities introduced. Results show that given a suitably reliable reference station, this method produces reliable adjustments to the mean, variance, and skewness.

1. Introduction

The ongoing debate about the natural and anthropogenic causes of climate change has made it important to investigate past climate for the longest possible time periods. However, all climate data have to be scrutinized thoroughly before they can be used to assess long-term changes in variability. Della-Marta et al. (2004) show that, on average, each of the 99 annual temperature records in Australia's high-quality dataset required 5–6 adjustments to the mean throughout the 100-yr record. These adjustments can be as large as a change in the climate that we are trying to detect and so without homogenizing the data, our conclusions could be seriously flawed. Australian climate data are certainly not the only data that suffer from homogeneity

problems. There are many studies that have either assessed the homogeneity (e.g., Manton et al. 2001; Serra et al. 2001; Wijngaard et al. 2003) or have homogenized climate series for use in long-term climate change assessment (e.g., Böhm et al. 2001; Tuomenvirta 2001; Camuffo 2002; Della-Marta et al. 2004; Auer et al. 2005; Begert et al. 2005). Many of these studies have focused on homogenizing monthly and annual average series, mainly because data are mostly available at that resolution and are readily accessible in digital form.

Often the climatic events that have the greatest impact on society are extreme events, and within the last 10 yr, there has been a greater emphasis placed on the research of these events (e.g., Trewin and Trevitt 1996; Collins et al. 2000; Trewin 2001; Jones and Lister 2002; Maugeri et al. 2002; Moberg et al. 2002; Vincent et al. 2002). Greater effort by National Meteorological Services (NMS), collaborative European Union projects (e.g., Camuffo and Jones 2002), and privately funded projects (e.g., Clarkson et al. 2001) has been put into digitizing daily data records in order to gain a greater

Corresponding author address: Dr. Paul Della-Marta, Climatology and Meteorology Research Group, University of Bern, Hallerstrasse 12, 3012 Bern, Switzerland.
E-mail: dmarta@giub.unibe.ch

understanding of climate variability and, most importantly, the nature and changes in the extreme parts of a climate variable's distribution. However, with this increase in the availability of daily data there has been a concomitant need for methods able to quality control and homogenize the higher-resolution data.

The methods used to homogenize annual and monthly climate data have often only adjusted the central tendency or mean state of the climate variable in order to obtain a long-term homogeneous series, which has been sufficient to provide reliable estimates of climate change and variability over many parts of the world (e.g., Houghton et al. 2001; Jones and Moberg 2003).

At the daily measurement resolution, it is likely that the typical inhomogeneities affecting a climate record are not adequately summarized by adjusting only the mean state, since we have evidence that these inhomogeneities and sometimes the processes that create them are nonlinear. From the study of micrometeorological processes (Oke 1987), it is evident that local climate can vary considerably on small space scales (~ 10 m) and that the processes that create these differences can only be modeled by complex dynamical equations or sometimes, at best, empirically. For example, experiments comparing modern instrument shelters with older ones usually show that the earlier shelters allowed the thermometers to be exposed to greater extremes of short- and longwave radiation (e.g., Parker 1994; Nordli et al. 1997) with nonlinear effects on climate throughout the seasonal cycle. In a statistical context, these nonlinearities are expressed as a change in the parent Cumulative Distribution Function (CDF), often with changes in higher-order moments of the CDF. Therefore, there is a need to make adjustments to the entire probability distribution when homogenizing daily data.

a. Previous studies on the homogenization of daily climate data

While the methods used to homogenize annual and monthly data are well established (e.g., Peterson et al. 1998; Ducre-Robitaille et al. 2003), relatively few methods exist to homogenize daily data.

Brandsma and Können (2006) present a technique called nearest neighbor resampling to homogenize daily mean temperature for inhomogeneities caused by a change in the time and frequency of subdaily measurements. They conclude that their method produces more reliable estimates of the changes in the 5th and 95th percentiles of mean daily temperature.

Vincent et al. (2002) use a simple approach to homogenize Canadian daily maximum and minimum temperature data based on adjustments identified using

standard techniques on monthly resolution data applied to the daily data. Adjustments to each daily temperature are made by fitting a piecewise linear function between the monthly mean adjustments such that the integrated magnitude of the daily adjustments preserves the monthly adjustments. They show that their method results in improved daily error estimates and greater spatial representation of extreme temperature trends. A similar technique has also been used by Maugeri et al. (2002, 2004) in the homogenization of daily temperature from Milan and daily sea level pressure data for six stations in the Po plain in Italy, respectively. They use a trigonometric function to interpolate daily corrections between the monthly corrections determined from relative homogeneity tests. Jones and Lister (2002) adjust daily mean temperatures from St. Petersburg, Russia, using differences between the monthly average of the raw daily values and the corrected monthly values from previous studies. Demarée et al. (2002) and Cocheo and Camuffo (2002) use overlapping records and linear models to adjust for screen bias and station relocations in order to construct a long-term central Belgium and Padova, Italy, temperature series, respectively. Moberg et al. (2002) apply daily corrections to the mean temperature series of Stockholm, Sweden, that had been linearly interpolated from monthly values. Their method also corrected a series for trend-type inhomogeneities caused by urban warming. Unfortunately, none of these methods adjust the higher-order moments explicitly.

The work of Allen and DeGaetano (2000), although not specifically applied to daily temperature data, describes a technique to adjust the frequency of exceedences of an annual extreme threshold. They conclude that their method could be applied to daily data.

One of the most robust methods capable of adjusting the higher-order moments of daily temperature data is that of Trewin and Trevitt (1996; detailed below), which builds a transfer function for the entire CDF.

In general, the literature concerning suitable techniques to homogenize daily temperature data is limited. The essence of the problem is that daily temperature measurements vary on relatively small spatial scales [typical decorrelation scales are ~ 200 km (Jones and Trewin 2002); however, these scales are likely to be smaller in areas of complex topography] and are influenced by local processes that are complex and nonlinear, which can be difficult to capture using our conventional climatological networks. Another problem could be that suitable overlapping records do not exist or are not readily accessible. In summary, we believe that the only previous study to explicitly homogenize maximum

and minimum daily temperature in moments higher than the mean is that of Trewin and Trevitt (1996).

b. The Trewin and Trevitt (1996) methods

Trewin and Trevitt (1996) present three different methods to build a composite daily temperature series. Essential to the methods is the existence of simultaneous (in time) observations from the *new* and *old* observing system. These parallel measurements had been taken based on the recommendations of Karl et al. (1995), who suggest that a minimum of a 2-yr overlap between the new and old observing systems be made. In Australia, for example, this practice has only become routine since around 1993 and so many inhomogeneities needed to be adjusted using the traditional constant difference techniques with neighboring reference stations. In this way, Trewin (2001) created a homogenized daily temperature dataset that has subsequently been used by Collins et al. (2000) to assess trends in the frequency of extreme temperature events in Australia.

The three methods they intercompared were constant difference, linear regression, and frequency distribution matching. The constant difference approach simply adjusted the older data with the newer data using the mean of the daily differences in the simultaneous (parallel) measurements. The linear regression method fitted a linear model to the difference in daily simultaneous measurements between the two observing systems and the temperature at the older station. This model could then be used to adjust daily temperatures at the older station differentially depending on the temperature, thereby adjusting the higher-order moments. Their third method determines the frequency distribution of each site during the simultaneous measurement period. The adjustment for each desired percentile is calculated simply as the difference between the two percentiles. This method assumes that there is no systematic bias in the rank order of the temperatures at the two sites. They show that both the regression method and the frequency distribution matching technique have certain advantages; however, if the homogenization of extreme events is most needed, then their frequency distribution matching technique is more accurate.

2. Data

The daily data used in this study consist of maximum temperature data from Graz University (Graz-Uni), Austria, and surrounding stations and artificial data. The annual daily data have been split into monthly series to avoid the treatment of the annual cycle. The

artificial data have been used to show the practicality of the method when the true solution to the inhomogeneity problem is known a priori.

To create the artificial daily candidate data, T , we started with Gaussian noise, Z_t , with a mean of zero and a variance (σ^2) of 1°C [$Z_t = \mathcal{N}(0, 1)$]. We then made it autocorrelated up to a lag of 6 days, which is typical of daily temperature data according to Eq. (1):

$$T_t = \alpha_1 T_{t-1} + \alpha_2 T_{t-2} + \dots + \alpha_6 T_{t-6} + Z_t, \quad (1)$$

where the autocorrelation coefficients $\alpha_{i=1,6} = (0.8, 0.7, 0.4, 0.3, 0.2, 0.1)$. We added a monthly mean and a variance inflation term to make winter more variable than summer months. As we will see later in section 4, the assumption that monthly daily temperatures are normally distributed is valid for our real example station detailed below. This we used as justification for creating the artificial data based on Gaussian noise; however, we note that often the assumption of normally distributed daily values is not correct. We also added a 1°C per 100 yr positive trend to it—this being typical of the magnitude for many stations around the world over the last century (Jones and Moberg 2003). To create the 10 reference series, we added Gaussian noise to the artificial candidate data with a standard deviation equal to half the monthly standard deviation of the candidate daily data. This resulted in daily series that were correlated between 0.8 and 0.9. We then added five randomly distributed inhomogeneities (within consecutive 30-yr periods) in the artificial candidate and reference series. All of the inhomogeneities in the mean were randomly created from a $\mathcal{N}(0, 1)$ distribution except for homogeneous subperiod (HSP) 4, where the change in skewness induced a change in the mean. Two of the inhomogeneities also included a change in the variance. This created six HSPs in which the data could be considered homogeneous (between the inhomogeneities). Table 1 summarizes the HSPs of the artificial candidate data and the inhomogeneities introduced.

The data used in section 3e relied on the creation of 500 reference series. These series were created similarly to the other artificial reference stations (detailed above); however, no break points in these series were introduced.

To test the method on real daily temperature data, we used maximum temperatures from Graz-Uni for the month of February. We also used seven other daily maximum temperature reference station records. The basic metadata of the seven stations and details of the homogeneous subperiods for the candidate and two reference stations believed to be the most suitable are

TABLE 1. A summary of the inhomogeneities in the artificial candidate series where T_t is the daily temperature at time t , μ is the mean of the artificial series, and αt is the trend term.

HSP	Time	Mean ($^{\circ}\text{C}$)	Higher-order moment factor
1	1979–2003	—	—
2	1960–78	−0.22	$= \begin{cases} 1.5T_t(T_t \leq \mu + \alpha t) \\ 1.5T_t(T_t > \mu + \alpha t) \end{cases}$
3	1915–59	+1.30	—
4	1889–1914	−0.52	$= \begin{cases} 1.5T_t(T_t \leq \mu + \alpha t) \\ 1.0T_t(T_t > \mu + \alpha t) \end{cases}$
5	1863–88	+0.77	$= \begin{cases} 0.5T_t(T_t \leq \mu + \alpha t) \\ 0.5T_t(T_t > \mu + \alpha t) \end{cases}$
6	1850–62	−0.84	—

compiled by Auer et al. (2001) and Böhm et al. (2001) (Tables 4, 5).

3. The higher-order moments method of homogenizing daily temperature

Where simultaneous observations between old and new observing systems do not exist or are not available, the most likely source of information regarding possible adjustments is neighboring stations. This concept has been utilized extensively in almost every other study on the homogenization of climate data and is also the basis of the new method we present here.

Although not treated in their paper, Trewin and Trevitt (1996) suggest that in order to build a daily composite temperature series that adequately adjusts the higher-order moments, another series from a nearby location (which is homogeneous) could be used when there are no overlapping measurements for the candidate station. We try to summarize the method as a sequence of operations defined in the following list:

- 1) Define HSPs for the *candidate* and as many *reference* stations as possible.
- 2) Starting with the most recent inhomogeneity, find a reference station that is highly correlated and has an HSP that adequately overlaps both HSP 1 and HSP 2 of the candidate station.
- 3) Model the relationship between the paired candidate and the reference observations before the inhomogeneity (i.e., in the period of common overlap within HSP 1 of the candidate).
- 4) Predict the temperature at the candidate station after the inhomogeneity using observations from the reference series in the period of common overlap within HSP 2 and the model.
- 5) Create a paired difference series between the pre-

dicted and observed temperature after the inhomogeneity (i.e., within HSP 2).

- 6) Find the probability distribution of the candidate station in HSP 1 and HSP 2.
- 7) Bin each temperature difference in the difference series (step 5) according to its associated predicted temperature in a decile of the probability distribution of the candidate station in HSP 1.
- 8) Fit a smoothly varying function between the binned decile differences (step 7) to obtain an estimated adjustment for each percentile.
- 9) Using the probability distribution of the candidate in HSP 2 (step 5), determine the percentile of each observation in HSP 2 and adjust it by the amount calculated in step 8.
- 10) HSP 2 is now homogeneous with respect to HSP 1. Repeat the process for all other HSPs sequentially.

The method described is more complex than those presented in Trewin and Trevitt (1996); however, it is similar conceptually. Both methods model the relationship between two series and both methods use estimates of the probability distribution of the candidate station. However, the higher-order moments (HOM) method is made more complex by the need to build a transfer function not just from one observing system to another (as is possible when parallel measurements are available), but a transfer function from one observing system (the candidate station before the inhomogeneity) to another observing system (a homogeneous reference station) to another observing system (the candidate station after the inhomogeneity).

We will now move on to a completed example of the method, and for this purpose we will use the artificial candidate data described in section 2 and hereafter refer to these as the candidate. Throughout the example and the remainder of the text, we will refer to the steps in the method defined above to help the reader. Using simulated data, we have defined the HSPs necessary in step 1 of the method.

a. Fitting the nonlinear model

By using a nonlinear model (step 3), it was believed possible to define some of the nonlinearities known to exist (Trewin and Trevitt 1996) in the relationship between two daily temperature records. Working backward from the latest inhomogeneity identified in the candidate (1978; see Table 1), we first find a highly correlated reference station (reference) with an HSP that spans at least 3 yr before (approximately 90 daily observations) and 3 yr after the candidate inhomogeneity (step 2). In our case, we used a reference that had a HSP from 1958–98. We fitted a nonlinear local re-

gression (LOESS; Cleveland and Devlin 1988) model to estimate the relationship between the candidate (response, y_i) and the reference (predictor, x_i) before the inhomogeneity (step 3). The smoothing model is given by Eq. (2),

$$y_i = g(x_i) + \epsilon_i, \quad (2)$$

where g is the regression surface, i is the i th observation from 1 to N_{model} (the total number of observations), and the ϵ_i are random errors. There are a number of parameters that control the regression surface g . They are the smoothing parameter, α ($\alpha > 0$), the degree of the local fitted polynomial λ (0, 1 or 2), and the distribution of the random errors, ϵ_i , either Gaussian or uniformly distributed. We chose the Gaussian distribution for the random errors since this has the effect of reducing the weight given to scatter points farther away from the fitted curve. The smoothing parameter and the degree of the polynomial fitted to the model was set so that the fitted function is smoothly varying and robust to outliers, especially at the extremes of temperature. The criteria for altering the parameters were set according to the correlation coefficient between the candidate and the reference, r , and the total number of observations ($N_{\text{total}} = N_{\text{model}} + N_{\text{pred}}$, where N_{pred} is the number of observations used to predict temperatures at the candidate after the inhomogeneity in HSP 2) used to build and apply the model specified below in Eq. (3):

$$\alpha = \begin{cases} 3.0, & (r < 0.8 \text{ } N_{\text{total}} < 1000) \\ 1.5, & (r < 0.8 \text{ } N_{\text{total}} \geq 1000) \\ 3.0, & (r \geq 0.8 \text{ } N_{\text{total}} < 1000) \\ 1.5, & (r \geq 0.8 \text{ } N_{\text{total}} \geq 1000) \end{cases}$$

$$\lambda = \begin{cases} 1, & (r < 0.8 \text{ } N_{\text{total}} < 1000) \\ 1, & (r < 0.8 \text{ } N_{\text{total}} \geq 1000) \\ 2, & (r \geq 0.8 \text{ } N_{\text{total}} < 1000) \\ 2, & (r \geq 0.8 \text{ } N_{\text{total}} \geq 1000) \end{cases} \quad (3)$$

A higher α results in a larger amount of smoothing, and the order of the polynomial, λ , controls the complexity of the fit. This precaution helped to prevent overfitting; however, initially the criteria were set in an arbitrary way from trial and error. Experimental evidence that the parameter settings are preventing overfitting is detailed in section 3e. In this case, a total of 620 daily values (N_{model}) were used to build the model. Figure 1 is a scatterplot of the candidate versus the reference and the LOESS-fitted function for HSP 1 (solid black line) and HSP 2 (dashed black line). Note that the model fitted to HSP 1 in Fig. 1 is the model we use in step 3 of the method and that the model fitted to

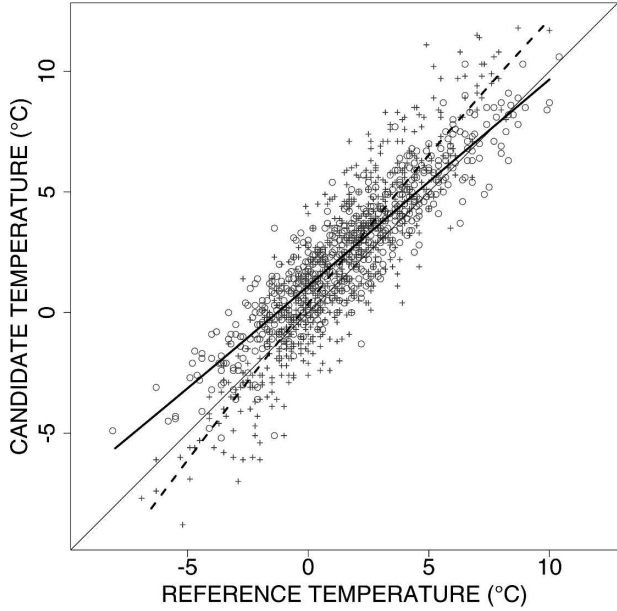


FIG. 1. The relationship between the artificial candidate station and the artificial reference series in HSP 1 (circles) and HSP 2 (crosses) where the solid and dashed curves denote the nonlinear LOESS fit, respectively. The thin black line has a slope of 1 for comparison purposes.

HSP 2 in Fig. 1 is displayed only for comparison purposes. The first thing to notice about the two fitted functions is that they are almost linear but with different slopes, and these slopes are both different from a slope of 1 (thin black line). The difference in the slope of the two fitted curves implies that a change in variance has occurred because of the inhomogeneity. We can also learn more about the variability of each station by comparing the slopes of the fitted curves to the line of slope equal to 1. In HSP 1, the solid black curve indicates that daily temperatures at the candidate are less variable than those at the reference, since its slope is less than 1. In HSP 2, the dashed black curve indicates that the variability at the candidate is more variable than at the reference, since the slope of the fitted curve is greater than 1.

b. Calculating the adjustments at each decile

The model (solid black line in Fig. 1; step 3) was then used to estimate the observations at the candidate after the inhomogeneity given homogeneous observations from the reference (using a total of 589 daily values, N_{pred} ; step 4). The differences between the candidate observed (inhomogeneous) values and the model-fitted values (step 5) after the inhomogeneity were binned (step 7) according to which decile the model-fitted values were placed in in the candidate observed CDF

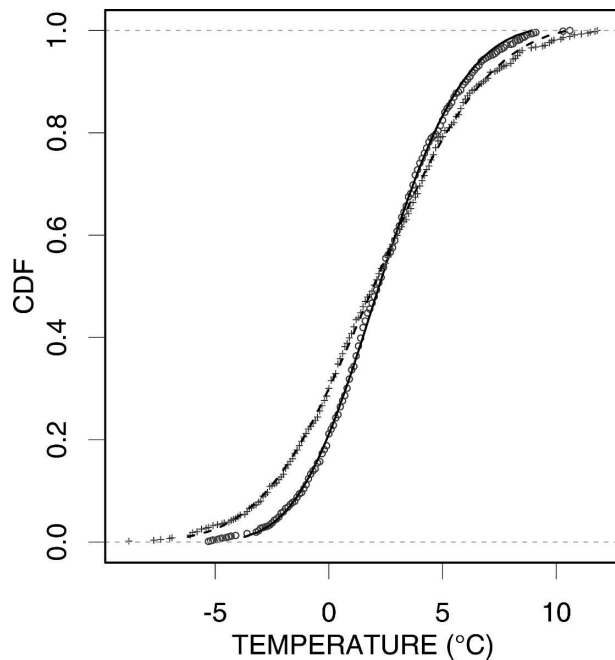


FIG. 2. The fitted (solid and dashed black line) and sampled (open gray circles and gray crosses) CDF of the artificial candidate station in HSP 1 and HSP 2 using the GEV distribution, respectively (see section 3 for details on the abbreviations of the fitted distributions).

(solid black line in Fig. 2; step 6), defined using the homogeneous temperatures before the inhomogeneity. By comparing the two CDFs in Fig. 2, it is apparent that the variance of the data in HSP 2 (dashed black line) is higher than in HSP 1 (solid black line). We fitted a LOESS function to the binned decile adjustments to obtain a smoothly varying set of adjustments between each decile, which are shown in Fig. 3 (step 8). Notice that the adjustment curve has an overall negative slope, which indicates that candidate data in HSP 2 should be made less variable to be homogeneous with HSP 1. The mean of the adjustments calculated over all deciles is $+0.5^{\circ}\text{C}$ and the known shift in the mean applied to the series was -0.2°C , showing that the adjustment curve is capturing both the change in variance and in this case, less accurately, the change in mean.

c. Applying the adjustments

Because the parent distributions of the candidate before and after the inhomogeneity are likely to be different (as indicated by the sample distribution shown in Fig. 2), it was necessary to apply the fitted adjustments (at each percentile) to the daily data according to the CDF of the candidate after the inhomogeneity (HSP 2; dashed black line in Fig. 2; step 9). This ensured that a

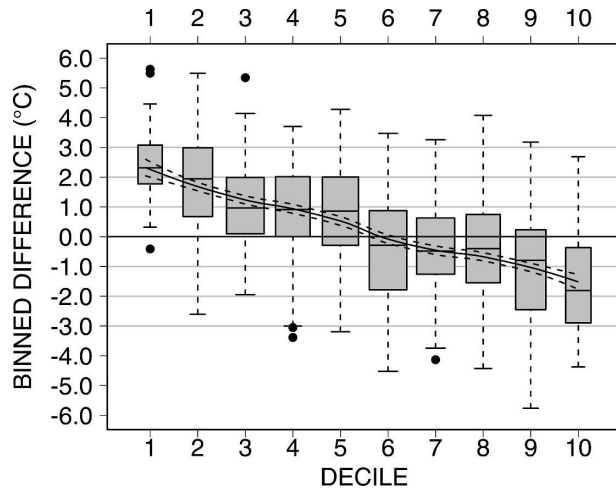


FIG. 3. The smoothed adjustments ($^{\circ}\text{C}$) for each decile shown as a solid black curve fitted using a LOESS function. The box plots indicate the mean of the binned differences (black line), the interquartile range (shaded area), 1.5 times the interquartile range (dashed black line), and outliers (dots). The width of the box indicates the relative number of observations in each. The dashed black curved lines show the 95% CI of the fitted curve.

daily value, say at the 95th percentile at the candidate after the inhomogeneity, was adjusted as if it were an observation at the 95th percentile in the candidate station in HSP 1, before the inhomogeneity.

In both cases, the CDF before and after the inhomogeneity was fitted using the theory of L-Moments (Hosking 1990) and the best fitting of six different distributions was estimated using a Kolmogorov–Smirnov test (Press et al. 1996). It is claimed by Hosking (1990) that L-Moments can provide more reliable estimates of the underlying distribution for small samples and are influenced less by outliers compared to conventional moments. The six distributions tested were the Normal (NOR), Generalized Extreme Value (GEV), Generalized Normal (GNO), Pearson type three (PE3), Generalized Pareto (GPA), and the Wakeby (WAK). For more information on the formulation of each distribution and the fitting method using L-Moments, readers are encouraged to refer to Hosking (1990). Generally, the NOR, GEV, GNO, PE3, and WAK distributions fitted the artificial data equally well before and after each inhomogeneity, which is not that surprising given that they are based on normally distributed data. However, the NOR distribution had a noticeably (not shown) poorer fit to the data in HSP 4, since it cannot model moments higher than the second.

Given the smooth nature of the sample CDFs (Figs. 2 and 4), it would have been possible to estimate the percentiles without fitting a model distribution to them. The HOM could be made simpler by using an empirical

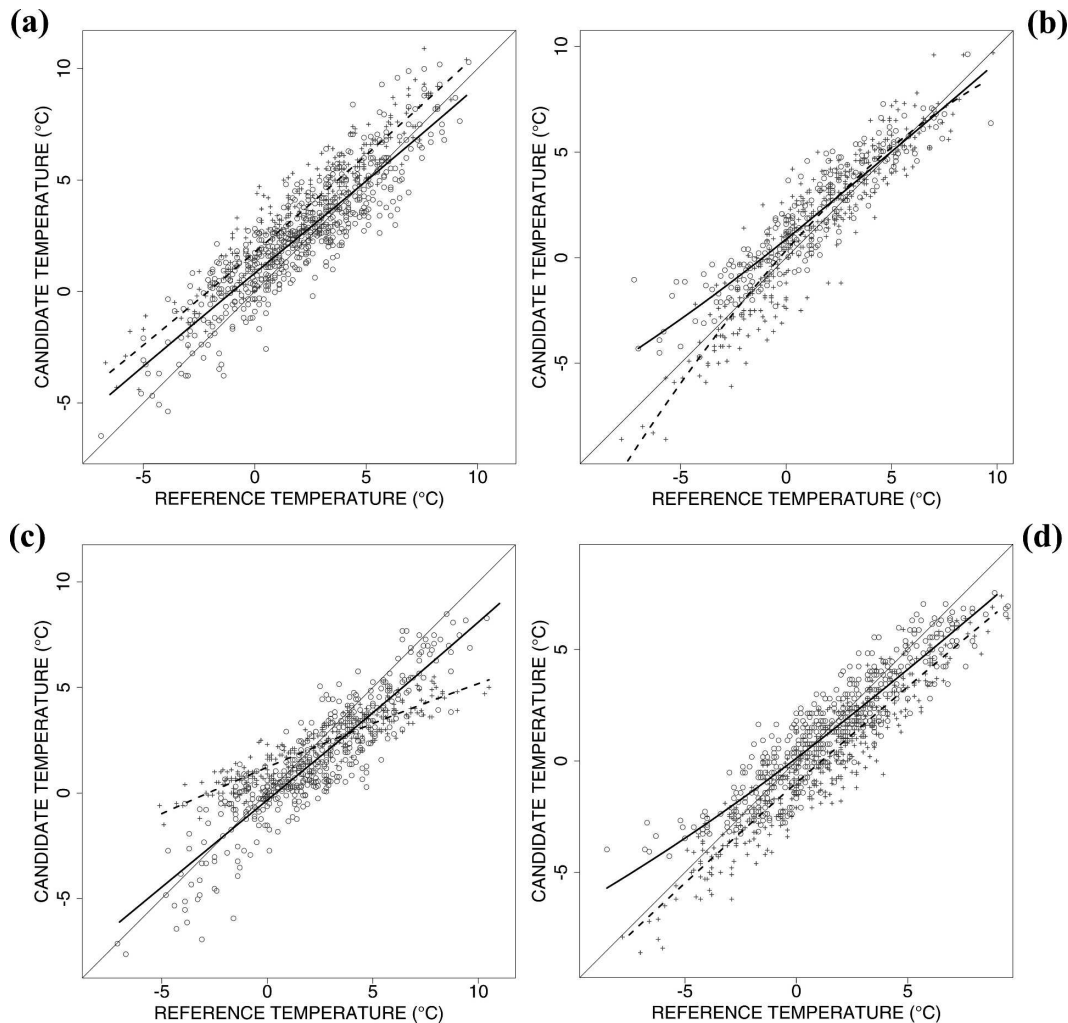


FIG. 4. The relationship between the artificial candidate station and an artificial reference series before (open gray circles) and after (gray crosses) with a LOESS-fitted curve before (solid black line) and after (dashed black line) each inhomogeneity in HSP 1 and (a) HSP 3, (b) HSP 4, (c) HSP 5, and (d) HSP 6. The thin black line in each plot has a slope of 1 for comparison purposes.

method to estimate the percentiles of the sample CDFs; however, a consequence of fitting a distribution using L-Moments is that it provides a useful estimate of the magnitude of the higher-order moments (not shown) that are not overly biased by outliers.

d. The other inhomogeneities

Now that we have applied the adjustments at each percentile, we have made HSP 2 homogeneous with respect to HSP 1. For simplicity in notation, we will now group what was HSP 1 and HSP 2 and call this HSP 1. The procedure of finding a suitable reference was repeated in order to adjust HSP 3 to be homogeneous with HSP 1 and so on, until the entire series was homogenized (step 10). The models, CDFs, and adjust-

ments for the other HSPs are shown in Figs. 4, 5, and 6, respectively. Concentrating on the models for the moment, we can see a wide variety of differences between the fitted curves. Figure 4a shows that the fitted curves are almost parallel with the dashed black line above the solid black line, indicating that temperatures in HSP 3 are on average higher than in HSP 1 (remember that HSP 1 is really the original HSP 1 and HSP 2 joined together after homogenization of HSP 2). This is also apparent in Fig. 5a; the dashed black line CDF is on average warmer than the solid black CDF. The adjustments show (Fig. 6a) that the method is capturing the known shift in the mean of $+1.3^{\circ}\text{C}$, with a mean negative adjustment of 1.0°C , although notice that the fitted curve to the adjustments has a slightly negative trend from decile 1 to decile 10. This is the result of a less-

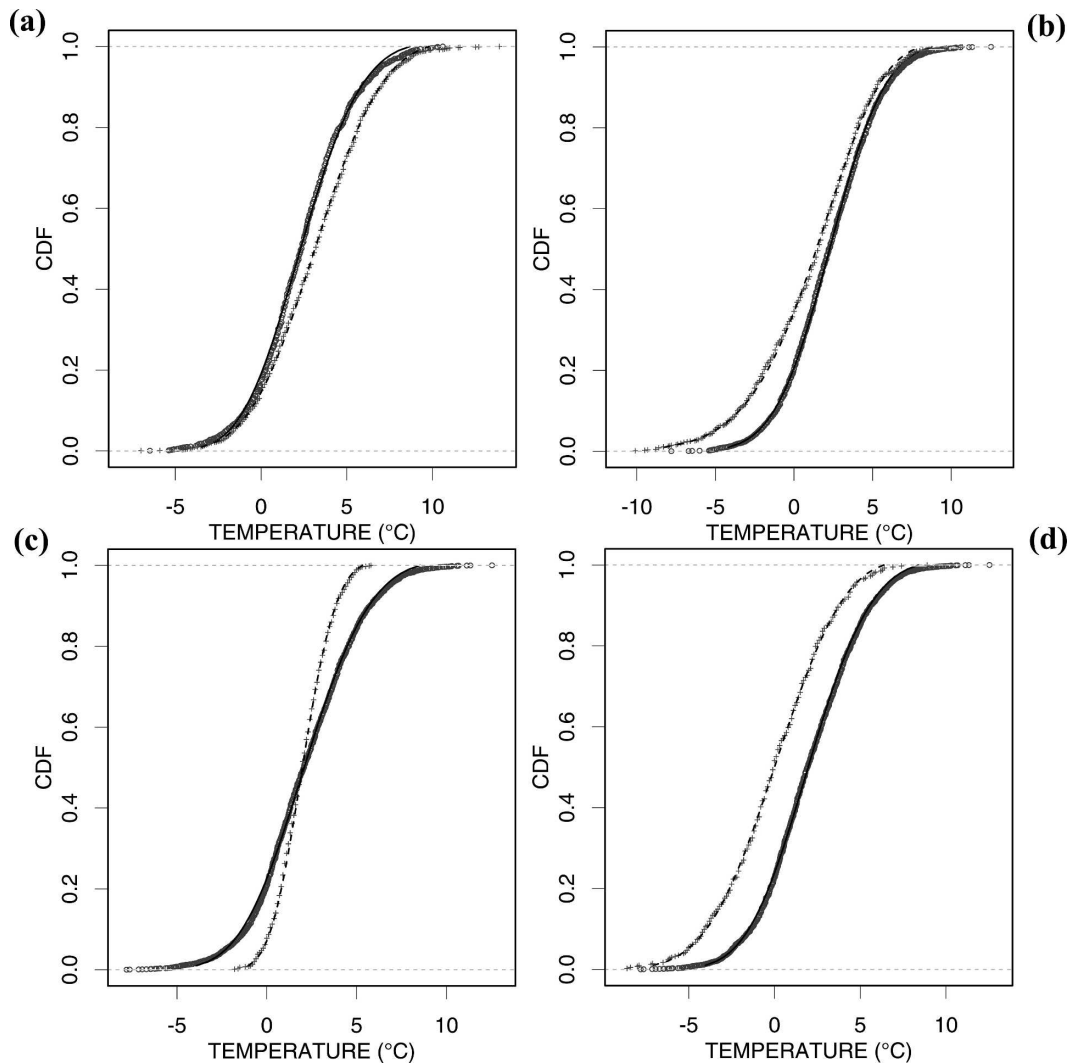


FIG. 5. The fitted and sampled CDF before (solid black line, open gray circles) and after (dashed black line, gray crosses) each inhomogeneity of the artificial candidate station: (a) HSP 1 and HSP 3 using the GEV distribution, (b) HSP 1 using the NOR and HSP 4 using the PE3 distribution, (c) HSP 1 and HSP 5 using the GEV distribution, and (d) HSP 1 using the PE3 and HSP 6 using the GEV distribution.

than-perfect model fit and sampling errors, since we know that the adjustments over all deciles should form a horizontal line. Similarly, the adjustments to HSP 6 (Fig. 6d) show mean adjustments of $+1.0^{\circ}\text{C}$ (actual inhomogeneity of -0.8°C). The fitted curve to the adjustments shows a small decreasing trend from decile 1 to decile 10. Again this seems to be the result of sampling errors and/or a poor model fit. If we look at Fig. 4d, we see that the reason for the small trend in the adjustments is an overfitting (solid black line) of the data in the lower decile.

In HSP 4, we applied a negative skewness that resulted in a true mean shift of -0.5°C . The fitted model and the adjustments show that this inhomogeneity was

captured well by our method. The dashed black line in Figs. 4b and 5b confirms that candidate has a negatively skewed relationship with reference in HSP 4 and is negatively skewed in comparison with candidate in HSP 1, respectively. The adjustments (Fig. 6b) show that candidate values less than the fifth decile in HSP 4 are adjusted increasingly positively toward the lower deciles. The mean adjustment is $+0.5^{\circ}\text{C}$ in agreement with the mean shift of -0.5°C .

The decreased variance applied to the candidate in HSP 5 was also well resolved by the method with an overall positive trend in adjustments from decile 1 to decile 10 (Fig. 6c) and with a mean adjustment of -0.6°C when the shift in the mean applied was $+0.8^{\circ}\text{C}$.

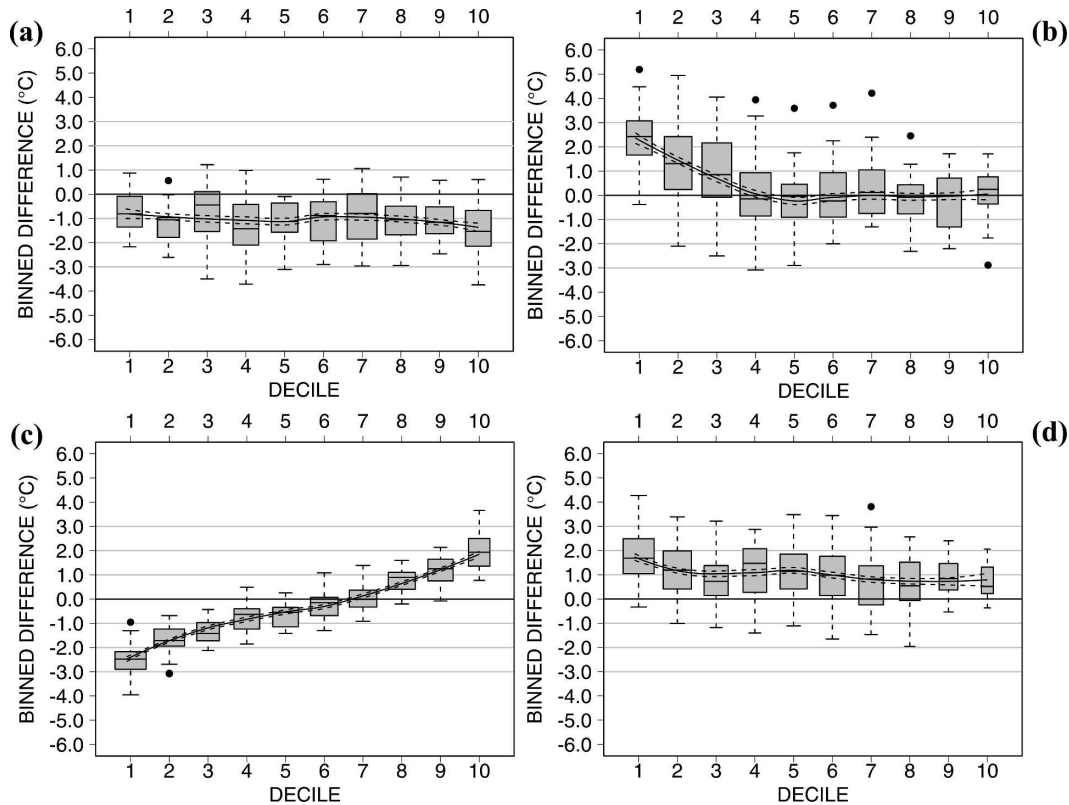


FIG. 6. Same as in Fig. 3, but for (a) HSP 3, (b) HSP 4, (c) HSP 5, and (d) HSP 6.

A comparison of the artificial monthly averaged daily data before and after homogenization is shown in Fig. 7. It is clear that the method is able to homogenize the data well compared to the truth (thin gray line); however, systematically the homogenized values tend to be higher than the true values in HSPs 2, 3, 5, and 6. This could be due to cumulative errors being propagated through the adjustments. This problem is common to most homogenization methods and is difficult to avoid; however, other examples of homogenized artificial data (not shown) compared with the true series show that the method is not systematically biased.

e. Experiments to assess the accuracy of the HOM method adjustments

In this section, we explore the sensitivity of the mean adjustment (calculated over all percentiles) and the overall root-mean-square error (RMSE) between the truth and the adjusted daily artificial series for various parameters of the nonlinear model fitted in steps 3 and 8. We will also show the results of altering other parameters such as the correlation between the candidate and the reference and the number of observations used to build (step 3) and apply (step 4) the model. All

experiments used the same artificial candidate series shown in Fig. 7 for an easy comparison of results. To build confidence intervals (CIs), we used 500 different reference series (see section 2 for the details). The parameters of each experiment are detailed in Table 2 and a summary of the results for each experiment (11 experiments in total, denoted as E1–E11) is shown in Table 3. The first eight experiments were designed to show the differences in the skill of the HOM method for each of the parameters specified in Eq. (3).

One of the most consistent results is the difference in μ_{RMSE} (see Table 3) between the experiments using a combined number of observations ($N_{\text{model}} + N_{\text{pred}}$) equal to 1000 and those equal to 200. Clearly the μ_{RMSE} are higher for those experiments using a lower number of observations. The number of observations also clearly makes a difference as to whether the HOM method is capable of reproducing the known shifts to the mean, μ_{true} . Looking at the results of E2 and E4, only the largest shift in the mean in HSP 3 is reliably adjusted when the number of combined observations is 200.

The effect of decreasing the correlation by increasing the noise added to the reference series, combined with

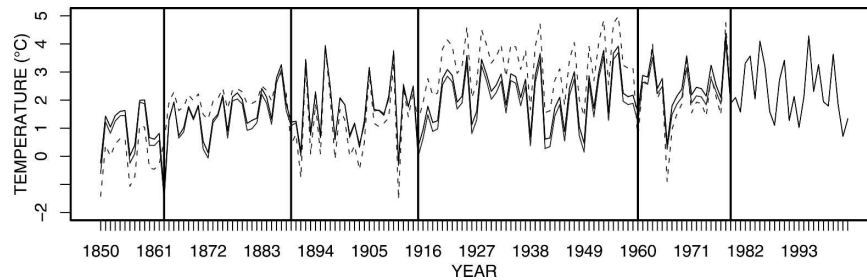


FIG. 7. A comparison of the monthly averaged daily artificial candidate station's true (thin gray line) and inhomogeneous (dashed gray line) time series, and the homogenized time series using the new method (solid black line). Black vertical lines denote the boundaries between the HSPs defined in Table 1.

a lower number of observations, results in unreliable adjustments for the HOMs and the mean. This is shown in the results of E6 and E8 where the correlation is around 0.7. Here, μ_{RMSE} (up to 2.4°C) are largest for the adjustments to HSP 3 (variance increased) and HSP 4, where a large negative skewness inhomogeneity was introduced. However, so long as a sufficiently large number of observations are used, the HOM method can reproduce the true series as shown in E5 and E7, where in every HSP (except HSP 2) the mean adjustment is confidently the same as the true adjustment (bold font in Table 3). The drawback of using lower-correlated stations is that the μ_{RMSE} values are sometimes doubled and generally increase from around 0.3°C in E1 and E3 to around 0.45°C in E5 and E7.

More subtle are the differences between the settings of α and λ used in steps 3 and 8 of the HOM method. There are noticeably better $\mu_{\text{HOMfitted}}$ results for E7

compared with E5 as the result of using a smoother LOESS fit (parameters shown in Table 2), shown by the lower 95% CIs in the former. Comparing the results of E1 and E3, we cannot see any systematic differences in the $\mu_{\text{HOMfitted}}$ or the μ_{RMSE} between the two experiments, indicating that the model is less sensitive to the overfitting when the correlation between the candidate and the reference station is high.

To test the hypothesis that a nonlinear model is better than a linear model used in steps 3 and 8, we designed E9 to be compared with E1 and E3. Instead of using the LOESS model [Eq. (2)], we used a simple least squares linear model. Overall, the linear model has given similar results to those in E1 and E3. The μ_{true} value is captured in all the $\mu_{\text{HOMfitted}}$ CIs. Also, the linear model produces similar μ_{RMSE} in all HSPs except HSP 4, where it is significantly greater by 0.2°C compared with the same HSP in E1 and E3. This indicates that a linear model could be used in steps 3 and 8 of the HOM method if there were no inhomogeneities in the third moment or skewness.

E10 was designed to test whether using lower values of $\alpha = 0.66$ than previously used in step 3 resulted in an overfitting of the model. Based on the results shown for E10, there is no evidence that overfitting is occurring. We have shown that in some circumstances overfitting is occurring (see black curve in Fig. 4d); however, it does not seem to overly affect the adjusted series statistics.

Finally, we wanted to test whether the parameters that control the fit to the binned decile differences (step 8) had a noticeable effect on the adjusted series statistics. In E11, we made the fit to the binned differences smoother; however, when comparing E11 to E1 and E3, there are no discernable differences.

A synthesis of these results suggests that the parameters chosen by trial and error in Eq. (3) to alter the sensitivity of the nonlinear model are a good approximation for the successful practical use of the method.

TABLE 2. Criteria of the various experiments performed to assess the accuracy of the adjustments to the synthetic candidate series detailed in section 2 and Table 1. Column 2 shows the proportion of the artificial candidate stations' standard deviation of white Gaussian noise added to the candidate station in order to create the reference series. Columns 3–6 show the parameters detailed in section 3.

Experiment number	Noise factor (reference)	N_{model}	N_{pred}	α, λ (step 3)	α, λ (step 8)
1	0.5	500	500	1.5, 2	0.66, 2
2	0.5	100	100	1.5, 2	0.66, 2
3	0.5	500	500	3.0, 1	0.66, 2
4	0.5	100	100	3.0, 1	0.66, 2
5	1.0	500	500	1.5, 2	0.66, 2
6	1.0	100	100	1.5, 2	0.66, 2
7	1.0	500	500	3.0, 1	0.66, 2
8	1.0	100	100	3.0, 1	0.66, 2
9	0.5	500	500	Linear	Linear
10	0.5	500	500	0.66, 2	0.66, 2
11	0.5	500	500	1.5, 2	1.5, 2

TABLE 3. Summary of the results of each experiment detailed in Table 2 using 500 different reference series for each candidate HSP. Column 2, μ_{true} , shows the mean adjustment applied to the HSP from Table 1. Column 3, r , shows the mean correlation between the candidate and the reference series in the period used to build the model (step 3) with corresponding 95% CI. Column 4, $\mu_{\text{HOMfitted}}$, shows the mean of the mean adjustment calculated over all percentiles with corresponding 95% CI (note that the adjustments in this column should be opposite in sign to the imposed shift in the mean shown in column 2). Column 5, μ_{RMSE} , shows the mean of the square root of the sum of daily errors between the truth and the HOM method adjusted series with a corresponding 95% CI. The $\mu_{\text{HOMfitted}}$ statistics that are statistically different from zero and capture the μ_{true} value are bold.

	μ_{true} (°C)	r	$\mu_{\text{HOMfitted}}$ (°C)	μ_{RMSE} (°C)
Experiment 1				
HSP 2	-0.22	0.92 ± 0.01	0.28 ± 0.21	0.32 ± 0.10
HSP 3	1.30	0.92 ± 0.01	-1.29 ± 0.19	0.23 ± 0.14
HSP 4	-0.52	0.91 ± 0.01	0.61 ± 0.27	0.31 ± 0.10
HSP 5	0.77	0.92 ± 0.01	-0.70 ± 0.21	0.36 ± 0.09
HSP 6	-0.84	0.92 ± 0.01	0.78 ± 0.26	0.27 ± 0.17
Experiment 2				
HSP 2	-0.22	0.92 ± 0.02	0.20 ± 0.44	0.60 ± 0.33
HSP 3	1.30	0.92 ± 0.02	-1.56 ± 0.54	0.69 ± 0.41
HSP 4	-0.52	0.92 ± 0.02	0.66 ± 0.75	0.79 ± 0.61
HSP 5	0.77	0.89 ± 0.06	-0.69 ± 0.70	0.66 ± 0.37
HSP 6	-0.84	0.91 ± 0.03	0.76 ± 0.85	0.52 ± 0.45
Experiment 3				
HSP 2	-0.22	0.92 ± 0.01	0.28 ± 0.19	0.32 ± 0.10
HSP 3	1.30	0.92 ± 0.01	-1.28 ± 0.20	0.22 ± 0.14
HSP 4	-0.52	0.91 ± 0.01	0.62 ± 0.26	0.30 ± 0.09
HSP 5	0.77	0.92 ± 0.01	-0.68 ± 0.20	0.35 ± 0.07
HSP 6	-0.84	0.92 ± 0.01	0.79 ± 0.24	0.25 ± 0.16
Experiment 4				
HSP 2	-0.22	0.92 ± 0.02	0.19 ± 0.47	0.57 ± 0.28
HSP 3	1.30	0.92 ± 0.02	-1.58 ± 0.54	0.69 ± 0.45
HSP 4	-0.52	0.92 ± 0.02	0.61 ± 0.75	0.71 ± 0.59
HSP 5	0.77	0.90 ± 0.06	-0.72 ± 0.69	0.64 ± 0.34
HSP 6	-0.84	0.91 ± 0.03	0.74 ± 0.80	0.57 ± 0.45
Experiment 5				
HSP 2	-0.22	0.76 ± 0.03	0.14 ± 0.31	0.44 ± 0.21
HSP 3	1.30	0.76 ± 0.03	-1.48 ± 0.29	0.39 ± 0.24
HSP 4	-0.52	0.73 ± 0.03	0.51 ± 0.39	0.48 ± 0.23
HSP 5	0.77	0.77 ± 0.03	-0.82 ± 0.33	0.50 ± 0.22
HSP 6	-0.84	0.77 ± 0.03	0.78 ± 0.47	0.51 ± 0.35
Experiment 6				
HSP 2	-0.22	0.77 ± 0.07	-0.04 ± 0.76	1.02 ± 0.86
HSP 3	1.30	0.77 ± 0.06	-2.19 ± 0.90	1.39 ± 0.96
HSP 4	-0.52	0.75 ± 0.07	0.52 ± 1.16	1.08 ± 0.81
HSP 5	0.77	0.72 ± 0.08	-0.62 ± 1.08	0.86 ± 0.51
HSP 6	-0.84	0.74 ± 0.08	1.04 ± 1.31	0.95 ± 0.80
Experiment 7				
HSP 2	-0.22	0.76 ± 0.03	0.16 ± 0.26	0.46 ± 0.15
HSP 3	1.30	0.76 ± 0.03	-1.46 ± 0.23	0.37 ± 0.22
HSP 4	-0.52	0.73 ± 0.04	0.53 ± 0.34	0.57 ± 0.13
HSP 5	0.77	0.76 ± 0.03	-0.78 ± 0.29	0.49 ± 0.16
HSP 6	-0.84	0.77 ± 0.03	0.82 ± 0.35	0.47 ± 0.30

TABLE 3. (Continued)

	μ_{true} (°C)	r	$\mu_{\text{HOMfitted}}$ (°C)	μ_{RMSE} (°C)
Experiment 8				
HSP 2	-0.22	0.77 ± 0.07	-0.06 ± 0.77	1.00 ± 0.73
HSP 3	1.30	0.78 ± 0.07	-2.21 ± 0.89	1.41 ± 0.88
HSP 4	-0.52	0.75 ± 0.07	0.49 ± 1.14	0.98 ± 0.71
HSP 5	0.77	0.72 ± 0.09	-0.65 ± 1.09	0.82 ± 0.53
HSP 6	-0.84	0.74 ± 0.08	0.99 ± 1.26	0.94 ± 0.77
Experiment 9				
HSP 2	-0.22	0.92 ± 0.01	0.23 ± 0.20	0.39 ± 0.07
HSP 3	1.30	0.92 ± 0.01	-1.32 ± 0.19	0.23 ± 0.15
HSP 4	-0.52	0.91 ± 0.01	0.63 ± 0.25	0.52 ± 0.06
HSP 5	0.77	0.92 ± 0.01	-0.66 ± 0.20	0.39 ± 0.05
HSP 6	-0.84	0.92 ± 0.01	0.81 ± 0.25	0.24 ± 0.19
Experiment 10				
HSP 2	-0.22	0.92 ± 0.01	0.27 ± 0.19	0.32 ± 0.10
HSP 3	1.30	0.92 ± 0.01	-1.30 ± 0.19	0.23 ± 0.15
HSP 4	-0.52	0.91 ± 0.01	0.61 ± 0.25	0.33 ± 0.10
HSP 5	0.77	0.92 ± 0.01	-0.70 ± 0.20	0.35 ± 0.09
HSP 6	-0.84	0.92 ± 0.01	0.78 ± 0.24	0.28 ± 0.18
Experiment 11				
HSP 2	-0.22	0.92 ± 0.01	0.24 ± 0.18	0.36 ± 0.08
HSP 3	1.30	0.92 ± 0.01	-1.32 ± 0.17	0.22 ± 0.14
HSP 4	-0.52	0.91 ± 0.01	0.57 ± 0.27	0.31 ± 0.09
HSP 5	0.77	0.92 ± 0.01	-0.71 ± 0.21	0.39 ± 0.08
HSP 6	-0.84	0.92 ± 0.01	0.76 ± 0.26	0.24 ± 0.18

4. Applying the method to Graz-Uni maximum temperatures

In this section, we show the practical application of the method to February daily maximum temperatures from Graz-Uni. Table 4 shows the HSPs identified by Böhm et al. (2001) and Auer et al. (2001) for the candidate and the reference series.

To homogenize HSP 1 with HSP 2, we used the HSP 1 of maximum temperatures from Vienna, Austria. The model (Fig. 8a) shows that before and after the inhomogeneity, there has been a general shift in the mean. There is also a subtle change in the relationship after the inhomogeneity in the upper tail of the distribution where it seems that the candidate is less extreme than it should be. This is also apparent by comparing the CDF of the candidate in HSP 1 and HSP 2 (Fig. 9a), where it is apparent that the area between the two CDFs is increasing above 5°C. Although the best fitting distributions were the WAK and GEV distributions for HSP 1 and HSP 2, respectively, other distributions including the NOR distribution gave an almost equally good fit. This was used as justification in section 2 to create artificial time series based on normally distributed data; however, we acknowledge that this is not always the case. The adjustments (Fig. 10a) indicate that a large

TABLE 4. A summary of the inhomogeneities in February maximum temperature series of Graz-Uni, Vienna, and Kremsmünster from Auer et al. (2001). The adjustments indicate the magnitude of the shift expressed as a cumulative adjustment.

Candidate			Reference stations					
Graz-Uni			Vienna			Kremsmünster		
HSP	Time	Mean °C	HSP	Time	Mean °C	HSP	Time	Mean °C
1	1988–2003	–	1	1953–2003	–	1	1986–2003	–
2	1955–87	+2.2	2	1902–52	+0.5	2	1965–85	+0.3
3	1943–54	+0.0	3	1900–01	+0.0	3	1946–64	+0.0
4	1891–1942	+2.0	4	1873–99	+0.4	4	1940–45	+0.1
5	–1890	+1.4	5	1853–72	–0.3	5	1909–39	–0.1
			6	–1852	+0.1	6	1891–1908	–1.0
						7	1887–90	–0.5
						8	1879–86	–0.3
						9	–1878	–0.2

shift in the mean of approximately 1.5°C is needed for deciles 1–4, but also a change in skewness is required by making the deciles 5–10 increase more extreme by up to 2.4°C. In general, these adjustments are similar in mag-

nitude to those suggested by Auer et al. (2001) and Böhm et al. (2001) shown in Table 4.

Using reference data from Kremsmünster’s HSP 3, the model and adjustments (Figs. 8b and 10b) indicate

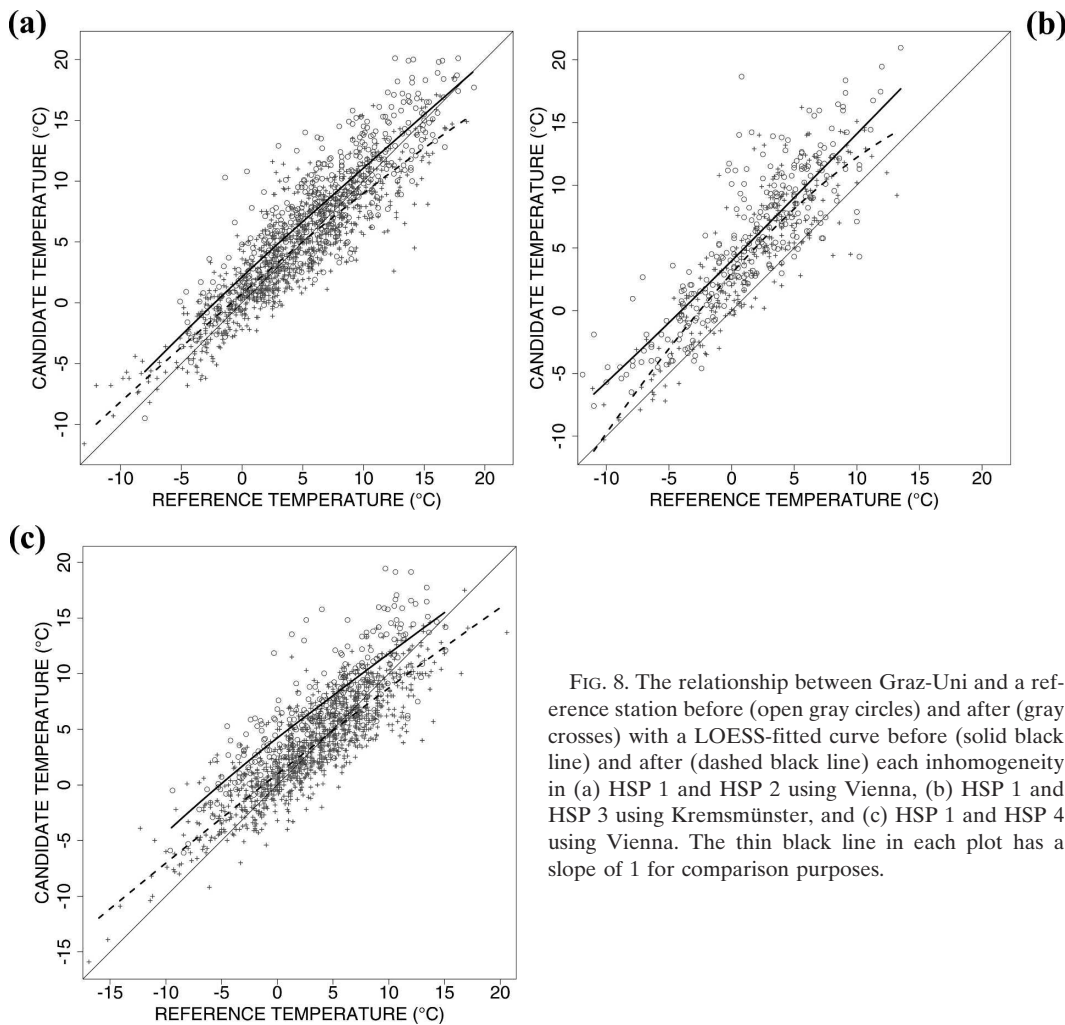


FIG. 8. The relationship between Graz-Uni and a reference station before (open gray circles) and after (gray crosses) with a LOESS-fitted curve before (solid black line) and after (dashed black line) each inhomogeneity in (a) HSP 1 and HSP 2 using Vienna, (b) HSP 1 and HSP 3 using Kremsmünster, and (c) HSP 1 and HSP 4 using Vienna. The thin black line in each plot has a slope of 1 for comparison purposes.

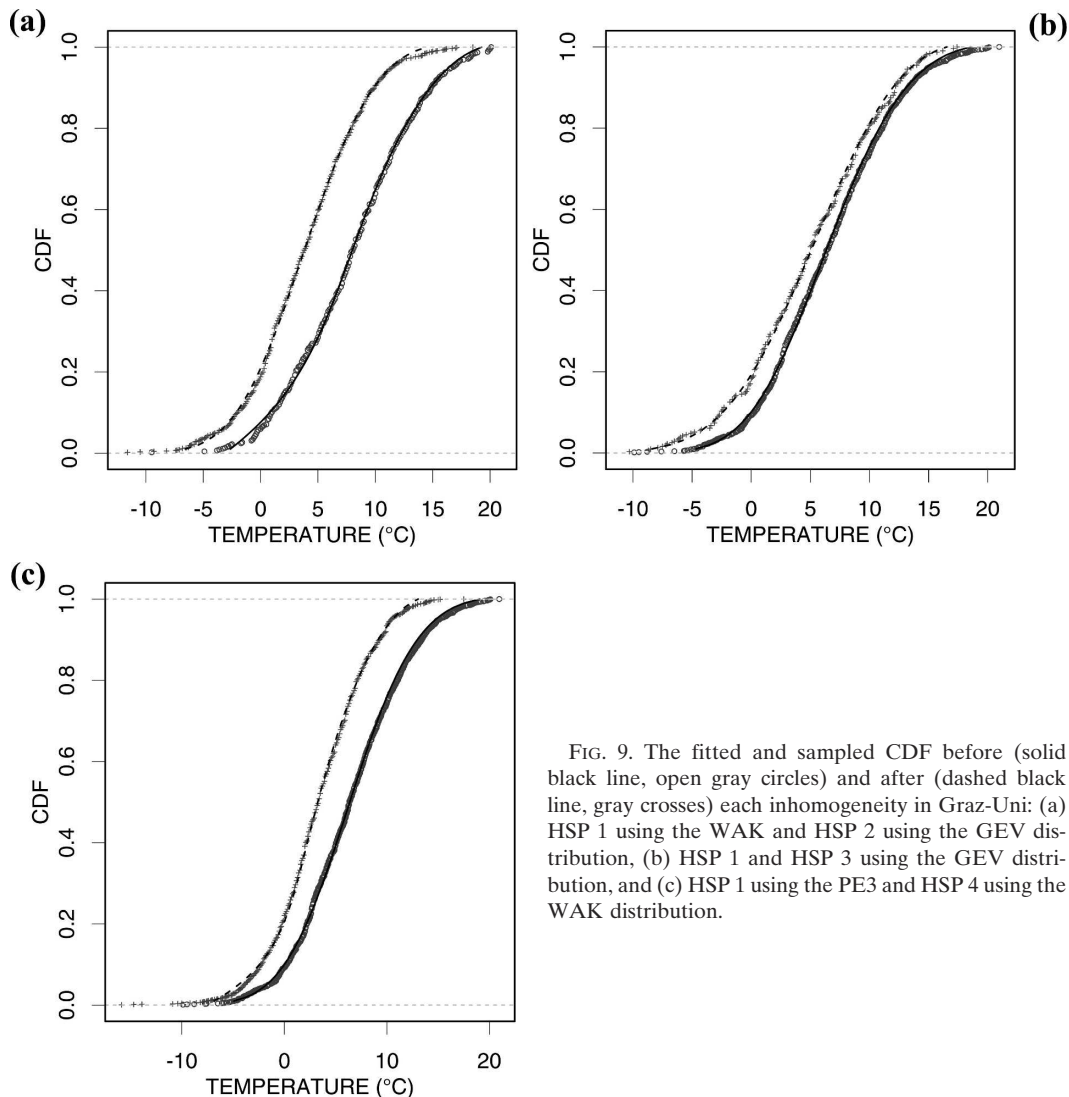


FIG. 9. The fitted and sampled CDF before (solid black line, open gray circles) and after (dashed black line, gray crosses) each inhomogeneity in Graz-Uni: (a) HSP 1 using the WAK and HSP 2 using the GEV distribution, (b) HSP 1 and HSP 3 using the GEV distribution, and (c) HSP 1 using the PE3 and HSP 4 using the WAK distribution.

that an increase in positive skewness and a large increase in the mean of approximately $+1.5^{\circ}\text{C}$ are necessary to homogenize it. This is substantially different from the negative adjustment (-2.2°C) suggested by Böhm et al. (2001) and Auer et al. (2001) that results in a cumulative adjustment of $+0.0^{\circ}\text{C}$ (Table 4). Since this finding is contradictory, it was necessary to investigate the possible causes more thoroughly. Detailed metadata for Graz-Uni were obtained from Auer et al. (2001) and indicate that generally, the Graz-Uni site has only experienced a few site relocations with no major relocations during the HSP 2 (1955–87) and HSP 3 (1943–54) under investigation. The cause of the inhomogeneity was most likely a change in the maximum temperature thermometer that seems to have occurred in late 1958 from an Amarell 850 to a Fuess 1106 type combined with a small (21 m) relocation of the screen

to a more open position (1946–87) from a position in close proximity to a wall (1.2 m away from the NNW facing wall, 1942–45). The instrument shelter was surely in a shaded position for the entire winter season during its close proximity to the NNW facing wall of the university, meaning that maximum temperatures were likely to have been cooler than when the screen was located in a more open position where it could have been exposed to periods of sunshine. By this reasoning, the temperature adjustments to the winter months should be positive as suggested by the HOM method. Another interesting piece of metadata is that the reference station Kremsmünster had a camouflage net cover during the years of 1943–45 and this could have been one reason why it was not included in the weighted reference series for Graz-Uni (I. Auer 2005, personal communication). This could have artificially

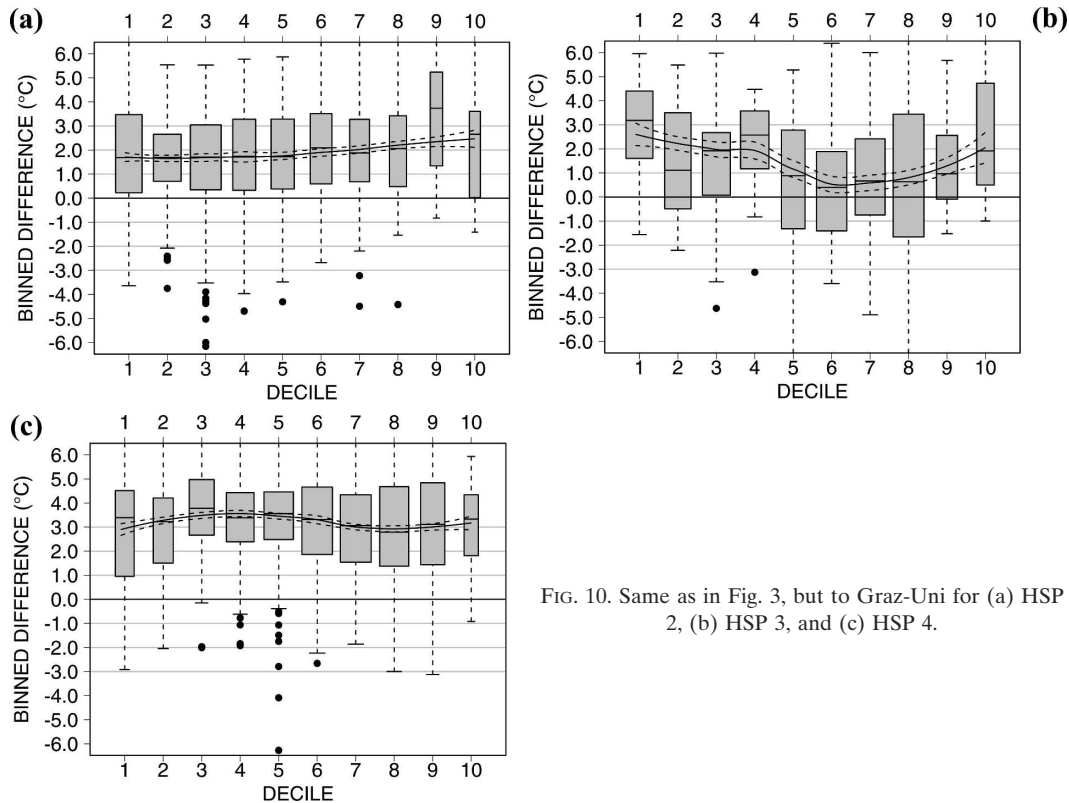


FIG. 10. Same as in Fig. 3, but to Graz-Uni for (a) HSP 2, (b) HSP 3, and (c) HSP 4.

influenced maximum temperature measurements and therefore altered the relationship between the stations. There could be a number of other reasons for this discrepancy, which are discussed in section 6. However, later in this section we present further evidence that the HOM method is producing an accurate result.

The adjustments to HSP 4 (Fig. 10c) were calculated based on the reference data from Vienna (HSP 2) and indicate that a large shift in the mean (averaged over all percentiles) of $+3.2^{\circ}\text{C}$ of the series is required. While this is in agreement with the suggested shift value of $+2.0^{\circ}\text{C}$, it is a significant difference. This large shift in the mean is also obvious from the fitted models shown in Fig. 8c (comparing before and after) and does not seem to be the result of model overfitting, since there is a high correlation between the two stations and a large number of observations used to build and apply the model.

Figure 11 compares February-averaged daily maximum temperature homogenized series using our method and the unhomogenized February averaged series. In general, it is apparent that the entire series before 1988 has been made warmer, which has substantially decreased the overall trend of February temperatures over the last 110 yr.

We extended our results and applied the HOM

method using a variety of reference stations to all other months of daily maximum temperature at Graz-Uni. The reference stations used above were the closest and most highly correlated, so we expect the HOM method to be most accurate using these stations. However, in order to assess the validity of the adjustments, we used a number of other surrounding stations that were not so well correlated to Graz-Uni (in no cases did we use a reference station listed in Table 5 that had a correlation of less than 0.75 with Graz-Uni).

Figure 12 compares the adjustments found by Auer et al. (2001) with the mean of the adjustments using the HOM method using four different reference stations for each HSP. For HSP 2 (Fig. 12a), it can be seen that there is good agreement in the sign and magnitude of the mean adjustments over all months; however, some notable exceptions exist using Zagreb, Croatia, as a reference station for February and March where differences are around 1.0°C . Generally, in HSP 4 (Fig. 12c), the sign of the mean adjustment was correct but the magnitude of the mean adjustment was larger using some reference stations rather than others. For example, the largest differences between the HOM method adjustments were in February using Vienna and Potsdam, Germany, as reference stations. Note that Potsdam is over 600 km away from Graz-Uni and

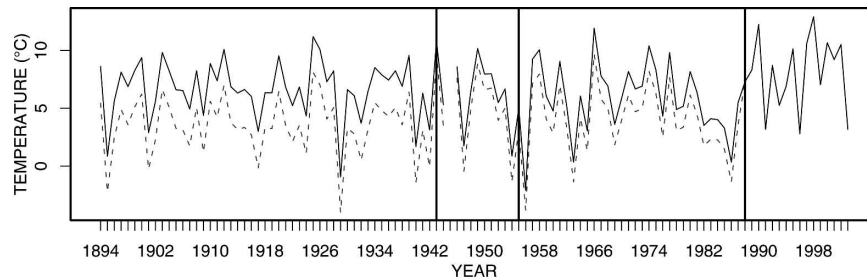


FIG. 11. A comparison of the February monthly averaged daily Graz-Uni inhomogeneous time series (dashed gray line) and the homogenized time series using the new method (solid black line). Black vertical lines denote the boundaries between the HSPs defined in Table 4.

so the confidence in this mean adjustment value is low. Other mean adjustments for February using Zagreb and Prague, Czech Republic, agree quite well. In HSP 3 (Fig. 12b), the mean adjustments agree more consistently in the summer half of the year than in the winter half. In most months, the HOM method adjustments are the same sign as the Auer et al. (2001) adjustments. The most striking of these differences is again the month of February. However, the fact that in the winter half of the year there is consistency between the mean adjustments using the HOM method at various reference stations indicates that the HOM method is accurate.

We also examined the percentile adjustments for each month and each HSP (not shown) for any systematic changes in the variance and skewness characteristics and conclude that for HSP 2 in the months of November, January, and February, there is an increase in variance required. All other months show a more constant adjustment over all deciles except May and October, which require a positive and negative skewness adjustment, respectively. The adjustments for HSP 3 are more varied and less systematic over the seasonal cycle; however, a positive skewness is needed in the months of February, March, May, and November and a decrease in variance in September, October, and De-

cember. For HSP 4, there was an increase in the variance of summer months needed whereas the remaining months showed a constant adjustment over all deciles.

5. The effects of using the HOM method on an extreme index

In this section, we compare the adjustments made to the artificial candidate station and Graz-Uni maximum temperatures using the HOM method and the more straightforward mean difference approach. We examined the effects of both methods on the variability of an extreme temperature index called txf_{90} , which is part of a family of indices commonly used in the assessment of changes to temperature extremes (e.g., Collins et al. 2000; Manton et al. 2001; Frich et al. 2002). The txf_{90} index simply counts the frequency of daily maximum temperatures above the long-term 90th percentile (in this case, the normal period used was the entire period), expressed as a percentage of days of a particular season or month that exceeds this threshold. The mean difference method was calculated using the HOM method as described in section 3, but instead of binning the differences between the fitted and observed series at the candidate station, we simply took the mean of the differences.

TABLE 5. Basic metadata of daily maximum temperature records used in this study ordered by the approximate distance from the candidate station, Graz-Uni. Details are obtained from Auer et al. (2001) for the Austrian stations and from the EMULATE (2002) database for all other stations.

Station name	Data availability	Latitude	Longitude	Elevation (m MSL)	Distance from Graz-Uni (≈km)
Graz-Uni	1894–2003	47°05′N	15°27′E	366	–
Vienna	1855–2003	48°14′N	16°21′E	198	144
Kremsmünster	1876–2003	48°03′N	14°08′E	383	146
Zagreb	1861–2003	45°49′N	15°58′E	157	146
Salzburg	1874–2003	47°48′N	13°00′E	437	200
Prague	1775–2003	50°04′N	14°25′E	191	340
Bamberg	1879–2003	49°52′N	10°52′E	282	458
Potsdam	1893–2003	52°22′N	13°04′E	81	611

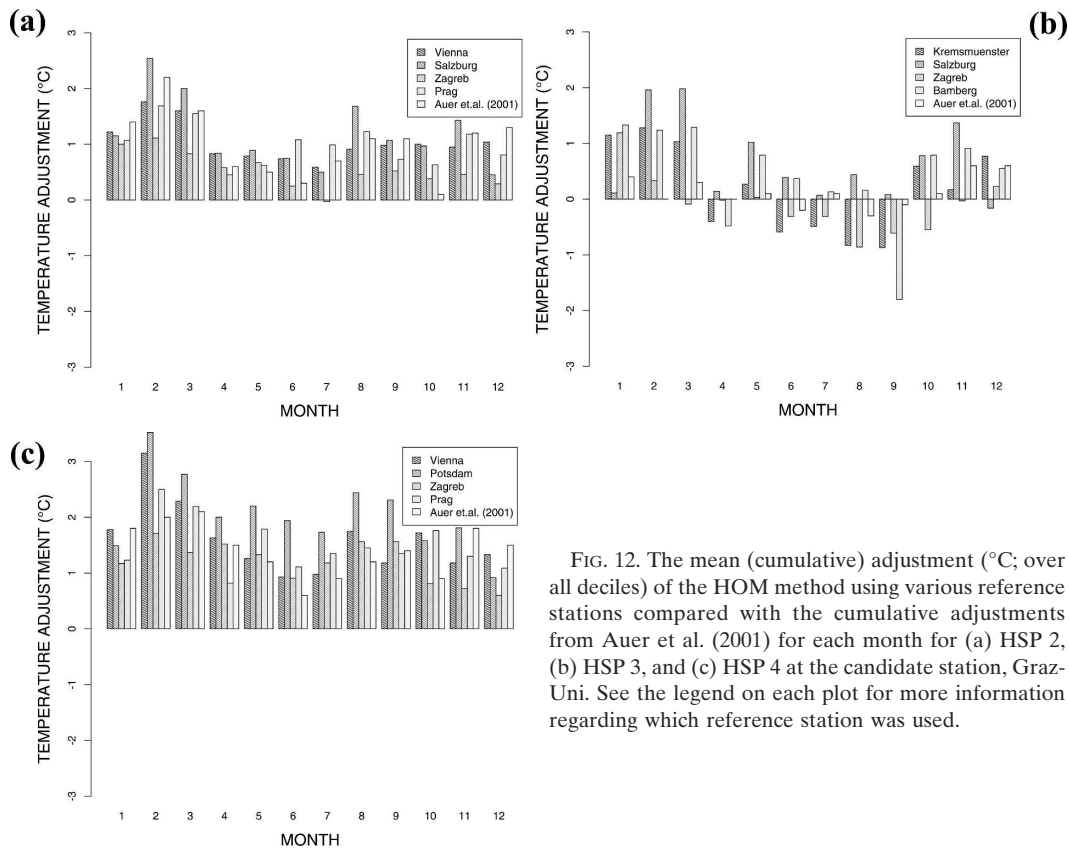


FIG. 12. The mean (cumulative) adjustment ($^{\circ}\text{C}$; over all deciles) of the HOM method using various reference stations compared with the cumulative adjustments from Auer et al. (2001) for each month for (a) HSP 2, (b) HSP 3, and (c) HSP 4 at the candidate station, Graz-Uni. See the legend on each plot for more information regarding which reference station was used.

Figure 13a shows that there is a considerable difference in the variability and the trend of txf90 calculated using the two methods on the artificial data. The index txf90 using the mean difference method shows large overestimations around 1970 in the artificial candidate series. Between 1863 and 1888, there are zero occurrences of txf90 in the series homogenized using the mean difference approach.

The differences between using the mean difference approach or the HOM method on the Graz-Uni txf90 series are not obvious from the linear trend (Fig. 13b) since they are almost equal; however, there are some differences in the variability of individual years. For instance, in the mid-1970s, the mean difference method underestimates txf90 by between 3% and 7%, while during some years after the 1990s, the mean difference method overestimates txf90.

6. Conclusions and discussion

We have shown the usefulness of a new method of homogenizing daily temperature measurements for changes in the CDF of a composite temperature series at moments higher than the mean.

When used on artificial data, the method reliably creates a daily composite record between two homogeneous subperiods of a candidate station that have different statistical moments of the mean, variance, and skewness. Typically, the magnitudes estimated at the 95% CI between the true inhomogeneity in the mean and the HOM method were within 0.3°C , even when there was a change in the HOMs as well. Further experiments with artificial data provided justification for using a nonlinear model instead of a linear model and for setting the parameters of the nonlinear model as we did. Typically, daily errors between the truth and the adjusted daily series were 0.2°C higher when using a linear model instead of a nonlinear model for HSP, which had a change in skewness. We showed that making the nonlinear model smoother resulted in better adjustment statistics when the correlation between the candidate and the reference was below 0.8.

When applied to real February daily maximum temperature data measured at Graz-Uni, Austria, there was a general agreement between the results of this study and previous studies with a significant exception for one inhomogeneity, which will be discussed in more detail below. The mean of the HOMs' method adjust-

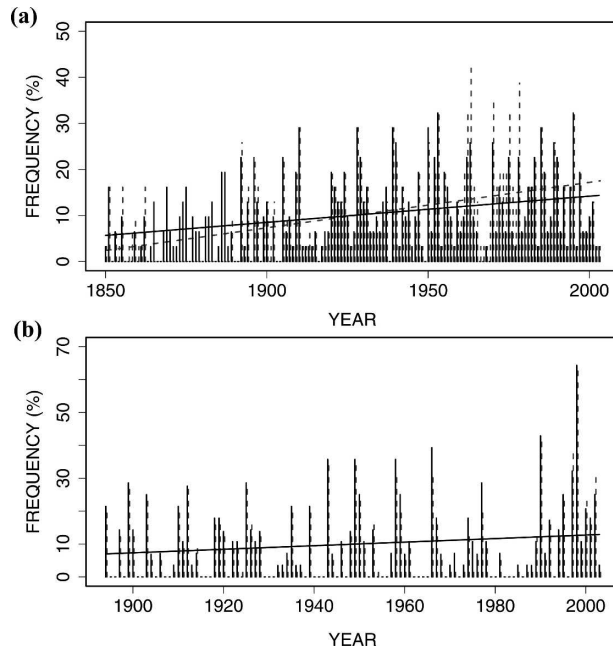


FIG. 13. A comparison of the derived monthly extreme indexes, tx90, and their linear trends with the raw data homogenized using the HOM method (black bars and line) and the mean difference method (dashed gray bars and lines) for the (a) synthetic candidate station and (b) February maximum temperature at Graz-Uni.

ments for the other months was shown to be coherent with the adjustments by Auer et al. (2001) in HSP 2 and 4, while in HSP 3, there was better agreement only in the summer months. However, using reference stations other than Vienna and Kremsmünster helped to confirm that the HOM method was giving consistent results. Closer examination of the adjustments over each decile and each month revealed that systematic changes in the variance and skewness characteristics were needed.

We propose that this method could also be used with daily minimum temperature, daily mean temperature series, and daily sea level pressure, since the spatial correlation is also as high as daily temperature or higher. However, this method is not likely to be suitable to homogenize daily precipitation measurements since the typical decorrelation scales are much lower, although it would be useful to test the method given a suitably dense network of stations (if they exist).

Previous methods of homogenizing daily temperature are not made redundant by this approach if the inhomogeneity is adequately modeled by a constant difference change in mean since this method can be sensitive to overfitting of the nonlinear model, especially when there is no obvious change in higher-order moments. However, as shown in section 5, if there is a significant change to the HOM of a candidate station,

the application of an adjustment to the mean only fails to adequately adjust the extremes and the derived extreme indices are not as accurate as they could be.

Our method could be used as an “add-on” homogenization procedure to many existing methods since it uses the HSPs identified by previous homogenization studies of monthly and annual temperature time series.

a. Some practical issues of using this method

To be as confident as possible in the adjustments, a suitable reference station must be chosen. In general, the reference station ideally has a high correlation with the candidate station. By this statement we mean that the candidate and reference stations should have a similar climate and be affected by the same local processes. As we have seen in section 3, if the correlation between the candidate and the reference was lower than 0.8 or if the number of observations used to build and apply the model were less than 1000, certain parameters of the model [detailed in Eq. (3)] were adjusted to make the model less sensitive to outliers or overfitting in the tails of the distribution. Our experiments in section 3e show that reference stations are ideally correlated at 0.9 or above with the candidate station; however, correlations with stations as low as 0.7 could still be useful given a sufficiently high number of observations to build and apply the model (greater than 1000). Since the method only uses one reference station to estimate the adjustments, it is important that all inhomogeneities in that series be accurately determined for the creation of the HSPs. If more than one suitable reference station is available, a comparison between the adjustments from each could be made. If the results were consistent then we may be more confident that adjustments to higher-order moments and the use of this method are justified. This was shown to be the case in section 4 where we applied four different reference stations to each HSP; however, we note that some of the reference stations were not ideal.

The remarks above lead to a discussion of the differences we observed between the suggested shift values for our example station, Graz-Uni. When we applied our method to February maximum temperature at Graz-Uni, there were some major differences in the adjustment applied to HSP 3 in the months of January–March. Based on the results in Fig. 12, it has been shown that the HOM method is giving congruent positive mean adjustments in the winter half of the year for HSP 3 that are generally larger than those shown in Auer et al. (2001). A more detailed comparison of our method to that of Auer et al. (2001) is needed to determine which of the two adjustments is more correct.

However, a result like this may not be unexpected

since Böhm et al. (2001), Auer et al. (2001), and many other previously mentioned homogenization studies have estimated their shifts to the mean from a difference series that has been created from a reference series that generally is a weighted average of many surrounding stations' monthly data. Weighted reference series generally have the advantage of being more representative of a climatic region, whereas a single reference station is representing the variability of climate on a smaller scale. A disadvantage of a weighted reference series could be that the individual station records used were not homogenized, so the series may not be truly representative of a larger area. The reference series can also have a different variance to the candidate series, which may influence the magnitude and significance of the mean shift value.

So, in summary, discrepancies like this will occur between the method we have presented here and previous studies based on a monthly weighted reference series because of the issue of spatial representation.

An implicit assumption of this method not previously mentioned is that the relationship between the candidate and the reference station is stationary over time. A change in this relationship can be viewed in two separate ways here. Either it changes because of the internal variability of a chaotic dynamical system or because of the fact that we had estimated the relationship using a finite sample of the parent distribution. Unfortunately, in this study it is not known if this assumption has been violated or not. Two ways to test the sampling variability would be to use a cross-validation approach to build the model and/or to use more reference stations where they are available.

b. Further research and possible improvements to the method

An obvious generalization of the method could be made to include many reference stations if they had suitable HSPs overlapping the candidate HSPs before and after an inhomogeneity. The spatial correlation structure of these surrounding reference stations could be utilized by a geostatistical method of interpolation to estimate the observations at the candidate station after the inhomogeneity. This may help address the issue of the spatial representativeness of the reference series.

Acknowledgments. We are grateful to Dr. Ingeborg Auer and Dr. Reinhard Böhm for the provision of metadata and inhomogeneity adjustments for the Austrian stations and for useful discussions relating to Austrian temperature series. We thank Dr. Olivier Mestre, Dr. Blair Trewin, and Jonas Bhend for helpful discussions and Dr. Jürg Luterbacher for revisions to the

manuscript and discussions. We would also like to thank two anonymous reviewers for their helpful comments and suggestions.

This research was funded by a European Union project called EMULATE (EVK2-CT-2001-00161). A contribution to the publication cost was made by the Marchese Francesco Medici del Vascello foundation.

REFERENCES

- Allen, R. J., and A. T. DeGaetano, 2000: A method to adjust long-term temperature extreme series for nonclimatic inhomogeneities. *J. Climate*, **13**, 3680–3695.
- Auer, I., R. Böhm, and W. Schöner, 2001: Austrian long-term climate 1767–2000: Multiple instrumental climate time series from central Europe. Zentralanstalt für Meteorologie und Geodynamik Tech. Rep. 25, Wien, Austria, 147 pp.
- , and Coauthors, 2005: A new instrumental precipitation dataset for the greater alpine region for the period 1800–2002. *Int. J. Climatol.*, **25**, 139–166.
- Begert, M., T. Schlegel, and W. Kirchhofer, 2005: Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000. *Int. J. Climatol.*, **25**, 65–80.
- Böhm, R., I. Auer, M. Brunetti, M. Maugeri, T. Nanni, and W. Schöner, 2001: Regional temperature variability in the European Alps: 1760–1998 from homogenized instrumental time series. *Int. J. Climatol.*, **21**, 1779–1801.
- Brandsma, T., and G. P. Können, 2006: Application of nearest-neighbor resampling techniques for homogenizing temperature records on a daily to sub-daily level. *Int. J. Climatol.*, **26**, 75–89.
- Camuffo, D., 2002: Calibration and instrumental errors in early measurements of air temperature. *Climatic Change*, **53**, 297–329.
- , and P. Jones, 2002: Improved understanding of past climatic variability from early daily European instrumental sources. *Climatic Change*, **53**, 1–4.
- Clarkson, N. M., B. C. Trewin, D. A. Jones, N. Plummer, R. L. Hutchinson, and W. Wong, 2001: Extending the computerised Australian climate archives to unlock our climate history—The CLIMARC project. *Proc. 14th Australia–New Zealand Climate Forum*, Darwin, Australia, Bureau of Meteorology, 49.
- Cleveland, W. S., and S. J. Devlin, 1988: Locally-weighted fitting: An approach to fitting analysis by local fitting. *J. Amer. Stat. Assoc.*, **83**, 596–610.
- Cocheo, C., and D. Camuffo, 2002: Corrections of systematic errors and data homogenisation in the daily temperature padova series (1725–1998). *Climatic Change*, **53**, 77–100.
- Collins, D. A., P. M. Della-Marta, N. Plummer, and B. C. Trewin, 2000: Trends in annual frequencies of extreme temperature events in Australia. *Aust. Meteor. Mag.*, **49**, 277–292.
- Della-Marta, P. M., D. Collins, and K. Braganza, 2004: Updating Australia's high-quality annual temperature dataset. *Aust. Meteor. Mag.*, **53**, 75–93.
- Demarée, G. R., P. J. Lachaert, T. Verhoeve, and E. Thoen, 2002: The long-term daily central Belgium temperature (cbt) series (1767–1998) and early instrumental meteorological observations in Belgium. *Climatic Change*, **53**, 269–293.
- Ducré-Robitaille, J. F., L. A. Vincent, and G. Boulet, 2003: Comparison of techniques for detection of discontinuities in temperature series. *Int. J. Climatol.*, **23**, 1087–1101.

- EMULATE, cited 2002: Emulate: European and North Atlantic daily to multidecadal climate variability. [Available online at <http://www.cru.uea.ac.uk/cru/projects/emulate/>.]
- Frich, P., L. V. Alexander, P. Della-Marta, B. Gleason, M. Haylock, A. M. G. Klein Tank, and T. Peterson, 2002: Observed coherent changes in climatic extremes during the second half of the twentieth century. *Climate Res.*, **19**, 193–212.
- Hosking, J. R. M., 1990: L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *J. Roy. Stat. Soc.*, **52B**, 105–124.
- Houghton, J. T., Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, and C. A. Johnson, Eds., 2001: *Climate Change 2001: The Scientific Basis*. Cambridge University Press, 881 pp.
- Jones, D. A., and B. C. Trewin, 2002: On the adequacy of digitised historical Australian daily temperature data for climate monitoring. *Aust. Meteor. Mag.*, **51**, 237–250.
- Jones, P. D., and D. H. Lister, 2002: The daily temperature record for St. Petersburg (1743–1996). *Climatic Change*, **53**, 253–267.
- , and A. Moberg, 2003: Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001. *J. Climate*, **16**, 206–223.
- Karl, T. R., and Coauthors, 1995: Critical issues for long-term climate monitoring. *Climatic Change*, **31**, 185–221.
- Manton, M. J., and Coauthors, 2001: Trends in extreme daily rainfall and temperature in southeast Asia and the South Pacific: 1961–1998. *Int. J. Climatol.*, **21**, 269–284.
- Maugeri, M., L. Buffoni, B. Delmonte, and A. Fassina, 2002: Daily Milan temperature and pressure series (1763–1998): Completing and homogenising the data. *Climatic Change*, **53**, 119–149.
- , M. Brunetti, F. Monti, and T. Nanni, 2004: Sea-level pressure variability in the PO plain (1765–2000) from homogenized daily secular records. *Int. J. Climatol.*, **24**, 437–455.
- Moberg, A., H. Bergstrom, J. R. Krigsman, and O. Svanered, 2002: Daily air temperature and pressure series for Stockholm (1756–1998). *Climatic Change*, **53**, 171–212.
- Nordli, P. O., H. Alexandersson, P. Frich, E. J. Forland, R. Heino, T. Jonsson, H. Tuomenvirta, and O. E. Tveito, 1997: The effect of radiation screens on Nordic time series of mean temperature. *Int. J. Climatol.*, **17**, 1667–1681.
- Oke, T. R., 1987: *Boundary Layer Climates*. 2d ed. Methuen, 435 pp.
- Parker, D. E., 1994: Effects of changing exposure of thermometers at land stations. *Int. J. Climatol.*, **14**, 1–31.
- Peterson, T. C., and Coauthors, 1998: Homogeneity adjustments of in situ atmospheric climate data: A review. *Int. J. Climatol.*, **18**, 1493–1517.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1996: *Numerical Recipes in FORTRAN 77: The Art of Scientific Computing*. 2d ed. Cambridge University Press, 933 pp.
- Serra, C., A. Burgueno, and X. Lana, 2001: Analysis of maximum and minimum daily temperatures recorded at Fabra Observatory (Barcelona, NE Spain) in the period 1917–1998. *Int. J. Climatol.*, **21**, 617–636.
- Trewin, B. C., 2001: Extreme temperature events in Australia. Ph.D. thesis, School of Earth Sciences, University of Melbourne, 239 pp.
- , and A. C. F. Trevitt, 1996: The development of composite temperature records. *Int. J. Climatol.*, **16**, 1227–1242.
- Tuomenvirta, H., 2001: Homogeneity adjustments of temperature and precipitation series—Finnish and Nordic data. *Int. J. Climatol.*, **21**, 495–506.
- Vincent, L. A., X. Zhang, B. R. Bonsal, and W. D. Hogg, 2002: Homogenization of daily temperatures over Canada. *J. Climate*, **15**, 1322–1334.
- Wijngaard, J. B., A. M. G. Klein Tank, and G. P. Konnen, 2003: Homogeneity of 20th century European daily temperature and precipitation series. *Int. J. Climatol.*, **23**, 679–692.